

Interaural correlation sensitivity

John F. Culling,^{a)} H. Steven Colburn, and Matthew Spurchise
*Department of Biomedical Engineering, Boston University, 44 Cummington Street,
Boston, Massachusetts 02215*

(Received 30 June 2000; revised 8 December 2000; accepted 3 May 2001)

Sensitivity to differences in interaural correlation was measured for 1.3-ERB-wide bands of noise using a 2IFC task at six frequencies: 250, 500, 750, 1000, 1250, and 1500 Hz. The sensitivity index, d' , was measured for discriminations between a number of fixed pairs of correlation values. Cumulative d' functions were derived for each frequency and condition. The d' for discriminating any two values of correlation may be recovered from the cumulative d' function by the difference between cumulative d' 's for these values. Two conditions were employed: the noisebands were either presented in isolation (narrow-band condition) or in the context of broad, contiguous flanking bands of correlated noise (fringed condition). The cumulative d' functions showed greater sensitivity to differences in correlation close to 1 than close to 0 at low frequencies, but this difference was less pronounced in the fringed condition. Also, a more linear relationship was observed when cumulative d' was plotted as a function of the equivalent signal-to-noise ratio (SNR) in dB for each correlation value, rather than directly against correlation. The equivalent SNR was the SNR at which the interaural correlation in an NoS π stimulus would equal the interaural correlation of the noise used in the experiment. The maximum cumulative d' declined above 750 Hz. This decline was steeper for the fringed than for the narrow-band condition. For the narrow-band condition, the total cumulative d' was variable across listeners. All cumulative d' functions were closely fitted using a simple two-parameter function. The complete data sets, averaged across listeners, from the fringed and narrow-band conditions were fitted using functions to describe the changes in these parameters over frequency, in order to produce an interpolated family of curves that describe sensitivity at frequencies between those tested. These curves predict the spectra recovered by the binaural system when complex sounds, such as speech, are masked by noise. © 2001 Acoustical Society of America. [DOI: 10.1121/1.1383296]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Pn [DWG]

I. INTRODUCTION

Several theories of binaural unmasking have emphasized the role of interaural correlation (Gabriel and Colburn, 1981; Durlach *et al.*, 1986; Koehnke *et al.*, 1986; Jain *et al.*, 1991; Culling and Summerfield, 1995; Bernstein and Trahiotis, 1992, 1996a, b). These authors have suggested that when an out-of-phase signal is added to a more intense, in-phase noise (NoS π) the interaural correlation of the stimulus is reduced at the signal frequency, and this reduction is detected by the listener and heard as a faint tone, or, if the noise is sufficiently narrow band, as a broadening of the sound image. As pointed out by Jain *et al.*, provided the signal is less intense than the masker, the reduction in correlation is monotonically related to the strength of the signal; the more intense the signal, the less the interaural correlation. This fact raises the possibility that listeners may be able to use interaural decorrelation as a reliable index of the signal intensity, as well as just for signal detection. Discrimination of the intensities of different components of a complex signal is often important in sound identification. In particular, speech recognition requires accurate estimation of the first formant frequency, which is vital for the identification of many

speech sounds. The estimation process is thought to be dependent upon the relative intensities of the group of peripherally-resolved harmonics in that frequency region, so an ability to discriminate different degrees of decorrelation would go some way towards explaining the binaural intelligibility level difference (Licklider, 1948; Carhart *et al.*, 1969a, b; Levitt and Rabiner, 1967a, b; Bronkhorst and Plomp, 1988).

Despite the importance of correlation discrimination, only a few papers have investigated listeners' ability to discriminate correlations (Pollack and Trittipoe, 1959a, b; Gabriel and Colburn, 1981; Grantham, 1982; Koehnke *et al.*, 1986; Jain *et al.*, 1991; Akeroyd and Summerfield, 1999), and all but one of these investigations reported discrimination for a reference correlation of 1.0 (and occasionally 0.0). The single exception is Pollack and Trittipoe's pair of articles, which measured correlation difference limens against seven reference values.¹ In their first article, these authors studied sensitivity to decorrelation for broadband sounds. They found that the difference limen was considerably smaller for correlations close to unity than for those close to zero. In their second article, they low- and high-pass filtered their stimuli in order to determine the role of different frequency regions. The latter study concluded that the most important frequency region was in the vicinity of 850 Hz for

^{a)} Author to whom correspondence should be addressed. Correspondence address: School of Psychology, Cardiff University, P.O. Box 901, Cardiff CF10 3YG, U.K. Electronic mail: cullingj@cardiff.ac.uk

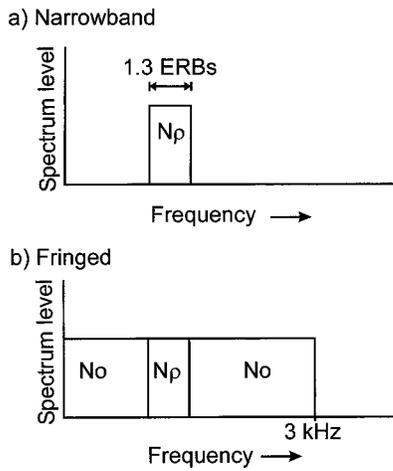


FIG. 1. Schematic illustration of the stimuli used in the two conditions.

a reference correlation of 0.0, but was 1700 Hz for a reference correlation of 0.5.

The present article aims to provide some primary data on interaural correlation sensitivity as a function of reference correlation and to encapsulate it into a simple usable form, so that it may be employed in predictions of the intensity difference limens of binaurally unmasked signals, binaural intelligibility level differences, dichotic-pitch percepts, and so forth. To this end, we have employed d' as a more flexible sensitivity measure than the difference limen and produced cumulative d' functions at each frequency. In an approach similar to that used by Baker *et al.* (1998) to describe the changing shape of auditory filters as a function of signal level and frequency, the data have been fitted with an eight-parameter function that controls the size and shape of the cumulative d' curve as a function of frequency.

II. EXPERIMENT

A. Stimuli

For each presentation interval, the 409.6-ms stimulus was generated immediately prior to its presentation using a TDT System 2 array processor at a 20-kHz sampling rate. Filtering was performed in the frequency domain, giving steep “brick-wall” cutoffs. All stimuli contained a target band that was 1.3 ERBs wide (Moore and Glasberg, 1983). The target band was centered at 250, 500, 750, 1000, 1250, or 1500 Hz. The corresponding bandwidths were, therefore,

68, 100, 133, 167, 202, and 238 Hz, respectively. The interaural correlation of the target band of noise was precisely controlled by mixing two orthogonal noises in appropriate ratios. Orthogonalization was achieved using the Gram–Schmidt procedure (Wozencraft and Jacobs, 1965), in which the correlated vector of two noises is subtracted from one of them and the waveform rescaled to the original amplitude (see the Appendix). Note that this contrasts with previous experiments in which randomly chosen samples of noise were combined; in those experiments the resulting correlation values in individual trials were randomly distributed about a chosen mean.

Two different conditions were investigated (see Fig. 1). In the narrow-band condition, the target band was presented in isolation. In the fringed condition, the target band was between spectrally contiguous flanking bands of noise (of the same spectrum level) which were also freshly generated for each trial. The flanking bands extended the spectrum of the stimulus and resulted in a flat spectrum between 0 and 3 kHz. The flanking bands had an interaural correlation of 1.0. The interaural correlations of the target bands depended on the condition. For each condition, a set of 15 pairs of correlation values were compared (Table I). In the narrow-band condition the values were clustered more closely than in the fringed condition at correlations close to 1.0. The differences in correlation were chosen (after pilot testing) to keep most of the d' values in the 0.5–2 range, which can be measured accurately without a very large number of trials. The inter-stimulus interval was determined by the processing time of the AP2. For the broadband condition, this was 1.2 s, while for the narrow-band condition it was 600 ms.

The sounds were presented to listeners in an IAC sound-attenuating chamber via a TDT System 2 rig (DD1 analog-to-digital converter; FT5-9 reconstruction filters; twin PA4 programmable attenuators; HB5 headphone amplifier) and Sennheiser HD414 headphones at an overall sound level of 87 dB(A) in the fringed condition. Sound levels were calibrated using a B&K artificial ear (type 4152), without flat-plate adapter, a B&K 1-in. microphone (type 4131) and B&K sound level meter (type 2203).

B. Procedure

Nine listeners took part in the study. Eight listeners, including the third author, completed both wideband and narrow-band conditions. Since five of these eight performed

TABLE I. The values of correlation, ρ , that listeners were required to discriminate in the fringed conditions and in the narrow-band condition.

		Fringed						Narrow band						
		Higher correlation						Higher correlation						
Lower correlation		0.3	0.5	0.65	0.8	0.9	1.0	Lower correlation	0.5	0.7	0.8	0.9	0.95	1.0
0.0		√	√	√				0.0	√	√	√			
0.3			√	√	√			0.5		√	√	√		
0.5				√	√	√		0.7			√	√	√	
0.65					√	√	√	0.8				√	√	√
0.8						√	√	0.9					√	√
0.9							√	0.95						√

relatively poorly in the narrow-band condition, the first author also contributed data in this condition. The listeners were advised at the start of the experiment to attend to different cues in different conditions and to use the cue resulting in best performance. In the case of narrow-band stimuli, listeners were advised to listen for changes in the breadth of the sound image within their heads. This advice was given because the narrow-band condition was similar to a narrow-band binaural unmasking experiment, in which the width of image is well known to be the optimal cue (Colburn, 1996, p. 344). In the fringed condition, they were advised to listen for a whistling sound (like a tone in a broadband binaural unmasking experiment) within the background noise which would be louder in one interval than in the other. In addition to this advice, listeners received trial-by-trial feedback throughout the experiment and during up to 20 h of practice prior to data collection.

Each d' value was measured for a given pair of correlation values (ρ_1 and ρ_2) by presenting listeners with a 55-trial series of 2I-FC comparisons. The results of the first five trials in each series were discarded. Each two-interval trial consisted of freshly generated examples of each of the two correlations under test. Since the fringed conditions required listeners to focus on a particular frequency region, the listeners were supplied with cueing sounds which directed their attention. These sounds were narrow-band stimuli at the same frequency with $\rho=1.0$ and they were presented once at the beginning of a series, then after the first five (discarded) trials and then every ten trials thereafter. During a 40- to 60-min session listeners completed all 15 series of comparisons for a given frequency and condition. The first five or six sessions in the fringed and narrow-band conditions were treated as practice for those conditions. Two sets of d' values were collected for each listener at each frequency and in each condition. In the first session for a given frequency and condition, the pairs of correlations used in each run had values that increased from one run to the next (from upper left to bottom right in Table I), while in the second session the sequence of runs was reversed. Each pair of $d'_{(\rho_1, \rho_2)}$ values was then averaged.

C. Results

1. Summarizing the data

Due to the large quantity of data collected, two different methods of summarizing data were developed. One method involved fitting values and the other functions to the raw d' data. Both used the "simplex" multi-parameter fitting procedure (Press *et al.*, 1988, Chap. 10.4) in order to fit cumulative d' values or functions directly to the raw d' measurements. Each fit assumed that listeners made use of a unidimensional decision axis for detecting correlation differences. Given this assumption (and equal variance), measured values of d' should be additive:

$$d'_{(\rho_1, \rho_3)} = d'_{(\rho_1, \rho_2)} + d'_{(\rho_2, \rho_3)} \quad (1)$$

where $\rho_1 \leq \rho_2 \leq \rho_3$.

The reader will note from Table I that, assuming that d' is cumulative as expected, there is some redundancy in the

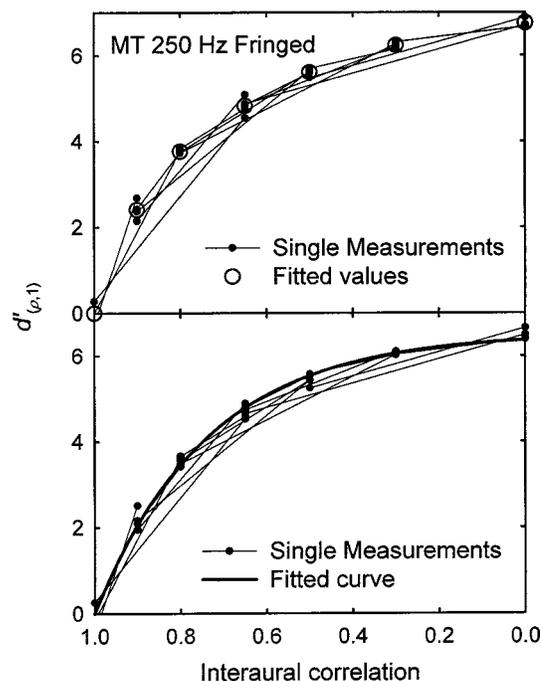


FIG. 2. Cumulative d' ($d'_{(\rho,1)}$) as a function of interaural correlation, ρ , for listener MT in the 250-Hz, fringed condition. The fitted $d'_{(\rho,1)}$ values are shown with large open circles (upper panel). The fitted function, created using Eq. (4), is shown with a thick line (lower panel). The individual measurements of d' taken from listener MT are shown on each panel using thin lines terminating in filled circles.

data collected. For instance, we collected $d'_{(0.3,0.5)}$, $d'_{(0.5,0.65)}$ and $d'_{(0.3,0.65)}$. If d' is cumulative, the last of these should approximate the sum of the first two. When the simplex procedure was used to fit the data, the fitted values and functions were those most consistent with all the available data. The procedure works by postulating values for a set of parameters, such as the parameters of a fitted function or specific $d'_{(\rho,1)}$ values, predicting the data based on these parameters, and then comparing the predicted with the observed data. The sum-of-squared errors, SS, between the predicted and observed data are used to evaluate the settings of the parameters. The parameter values are then permuted and the evaluation repeated. This cycle continues until SS has been minimized. It is the method of parameter permutation which makes the simplex procedure an efficient multi-parameter fitting algorithm (for further details, see Press *et al.*, 1988, Chap. 10.4).

2. Cumulative sensitivity values

The first fitting procedure was used to fit six cumulative d' values, $d'_{(\rho,1)}$, to each set of 15 measured $d'_{(\rho_1, \rho_2)}$ values collected from a given listener, in a given condition (fringed/narrow band) and at a given frequency. $d'_{(\rho,1)}$ was evaluated for each nonunity value of ρ (e.g., 0.0, 0.3, 0.5, 0.65, 0.8, and 0.9). The results of this fitting procedure are illustrated by the large open circles in Figs. 2 and 3. Ideally, the individually measured values of $d'_{(\rho_1, \rho_2)}$ and the values of $d'_{(\rho,1)}$ derived from the fitting process should all observe Eq. (2):

$$d'_{(\rho_1, \rho_2)} = d'_{(\rho_1, 1)} - d'_{(\rho_2, 1)} \quad (2)$$

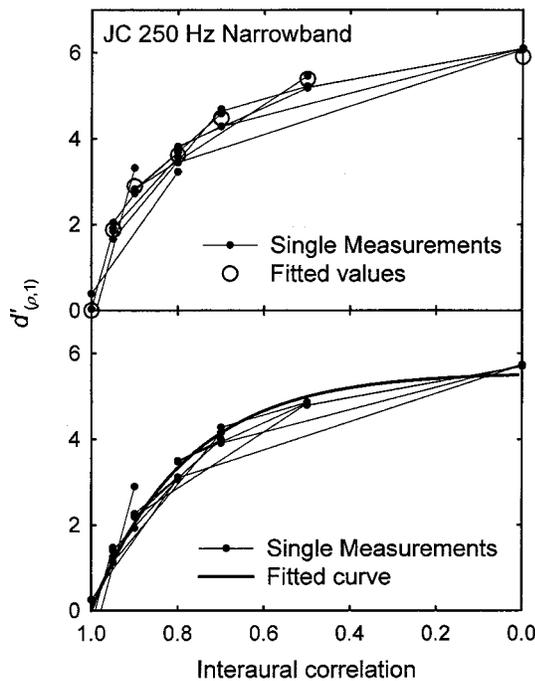


FIG. 3. As in Fig. 2, but for listener JC in the 250-Hz, narrow-band condition.

Deviation from this equivalence was assumed to represent a combination of measurement noise and suboptimal fitting of $d'_{(\rho,1)}$ values. Such deviation was therefore squared and summed across the 15 measured values of $d'_{(\rho_1, \rho_2)}$ to give the error term, SS, to be minimized by the fitting program.

The results of this fitting process can be seen in the large open symbols on the upper panels of Figs. 2 and 3. The thin lines terminated with filled circles represent the 15 $d'_{(\rho_1, \rho_2)}$ values to which the fit was made in each case. These lines are plotted between the appropriate correlation values on the abscissa, and their vertical extents correspond to the observed $d'_{(\rho_1, \rho_2)}$ values. Their vertical position minimizes the deviation between the filled symbols at each end and the corresponding open symbols. The total squared deviation across all these symbols therefore represents SS.

3. Correlation sensitivity as a function of correlation

The second fitting procedure was used to generate a summary of the form that appears in the thick curves in Figs. 2 and 3. These fits turned the same sets of 15 $d'_{(\rho_1, \rho_2)}$ values into curves describing cumulative d' as continuous functions. Continuous cumulative d' functions were fitted to the measured $d'_{(\rho_1, \rho_2)}$ values using the base equation:

$$d'_{(0,\rho)} = e^{(k\rho+n)} - e^n, \quad (3)$$

where k and n are fitted parameters. This equation was chosen because it always evaluates to zero at $\rho=0$, while the two parameters (in combination) control the rate of change of curvature and the value at $\rho=1$. Henceforth, an equivalent expression will be used for the change in correlation from $\rho=1$:

$$d'_{(\rho,1)} = e^{(k+n)} - e^{(k\rho+n)}. \quad (4)$$

The lower panels of Figs. 2 and 3 illustrate the quality of fit that is achieved using Eq. (4) for the fringed and narrow-band conditions, respectively. The thick lines are the fitted cumulative d' functions. Again, the fit minimized the squared errors between the observed $d'_{(\rho_1, \rho_2)}$ values and the differences between the $d'_{(\rho_1,1)}$ and $d'_{(\rho_2,1)}$ values from the fitted curve. The lower panels include the same thin lines terminated by filled symbols as the upper panels, but adjusted in vertical position, so that they reflect the quality fit achieved by the fitted functions.

As can be seen from Figs. 2 and 3, the quality of the fitted curve is acceptable. For the fit in Fig. 2, $SS=1.90$, slightly above the average value across all 17 data sets (1.51). In Fig. 3, an example of one of the poorest fits, $SS=5.19$. The high error is largely attributable to one unusually large d' measurement for the discrimination of correlations 0.9 and 1.0. Since the principle source of errors in the fits appeared to be noise in the measurements, Eq. (4) was adopted as a suitable fitting function to be used in the larger-scale fitting process to follow.

The functions in Figs. 2 and 3 are typical results for the dependence of $d'_{(\rho,1)}$ on $1-\rho$, the reduction in correlation from unity. These results show the characteristically higher sensitivity to changes in ρ near unity correlation and the reduction in sensitivity near zero correlation. Performance in the fringed and narrow-band conditions are similar at 250 Hz in these examples. However, there are substantial variations across frequencies, between these two conditions and across listeners.

4. Correlation sensitivity at different frequencies

The thick curves in Figs. 4 and 5 show further functions fitted [using Eq. (4)] to the data at a single frequency for listeners RM and MS at each frequency tested. The symbols in Figs. 4 and 5 show the results of the first fitting procedure. These symbols show $d'_{(\rho,1)}$ at each of the correlation values used in the experiments.

Note that listener RM in Fig. 4 performed poorly in the narrow-band condition compared to the fringed condition while for listener MS in Fig. 5 the reverse occurred. Cumulative d' values between $\rho=0$ and $\rho=1$ ($d'_{(0,1)}$) are provided for each listener at each frequency in both conditions in Table II. Since the first five listeners in the table seemed to perform worse than the last four in the narrow-band condition, we chose to characterize their data separately in subsequent analyses (the ordering of the listeners in Table II does not reflect the order in which they participated in the experiment).

In general, performance decreases above 750 Hz. In addition, for most cases (and for 20 of 24 comparisons at the three lowest frequencies), the fringed stimuli have higher $d'_{(0,1)}$ values than the equivalent narrow-band conditions. This difference is remarkable because the fringes carry no additional information and could be considered simply as masking the information in the target band. However, as discussed below, the task is subjectively easier for most listeners with the fringe present.

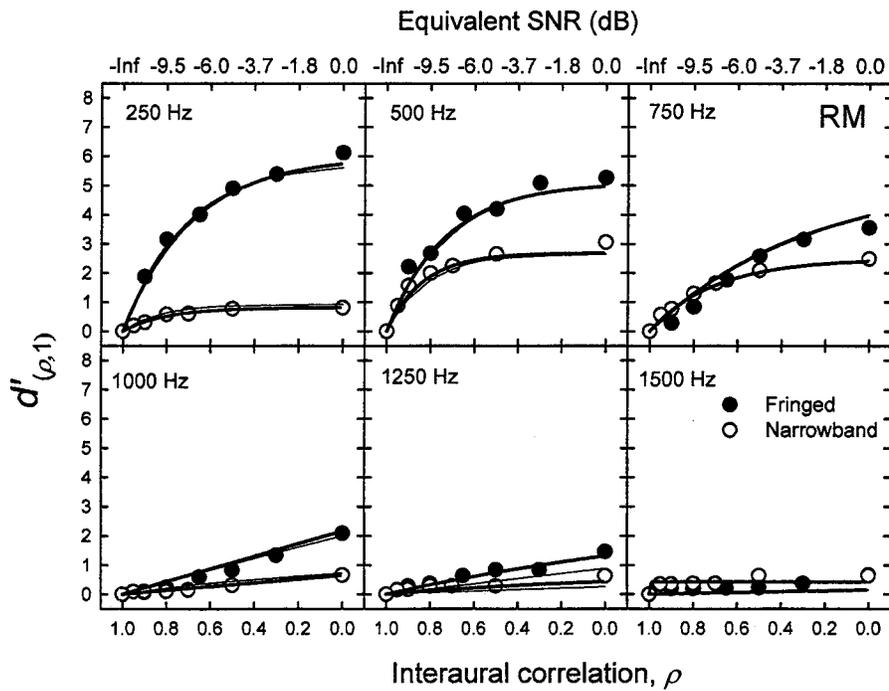


FIG. 4. Cumulative d' ($d'_{(\rho,1)}$) as a function of interaural correlation, ρ , at six frequencies for listener RM. The symbols are $d'_{(\rho,1)}$ values fitted to the raw data at each frequency. The thick lines are $d'_{(\rho,1)}$ functions fitted to the raw data at each frequency using Eq. (4). The thin lines are $d'_{(\rho,1)}$ functions fitted using the across-frequency fitting method based on Eqs. (5) and (6). The open circles and intersecting lines are for the narrow-band condition. The filled circles and intersecting lines are for the fringed condition. The top abscissa is marked with the signal-to-noise ratios that would produce the corresponding correlation values from the bottom abscissa if signal and noise were added in the NoS π binaural configuration.

5. Correlation sensitivity as a function of correlation and frequency

Having established that Eq. (4) is well suited to represent the growth of cumulative d' , it was also used to fit the data across different frequencies. Here, the parameters n and k were not fitted separately for each frequency. Instead, the data from all frequencies were fitted simultaneously using functions [Eqs. (5) and (6)] that related n and k to frequency,² f :

$$k = \frac{r_k}{1 + e^{s_k(f-t_k)}} + a_k, \quad (5)$$

$$n = \frac{r_n}{1 + e^{s_n(f-t_n)}} + a_n. \quad (6)$$

These functions are both logistic curves providing a sigmoidal transition between two asymptotic values. The parameters r , s , t , and a are free parameters of the fit: r controls the absolute range of the parameter, s controls the steepness of the transition in its value as a function of f , t controls the frequency of the transition, and a controls the asymptotic value as $f \rightarrow \infty$. The choice of logistic functions was motivated by the reduction in the binaural masking level difference (BMLD) and in correlation sensitivity at high frequen-

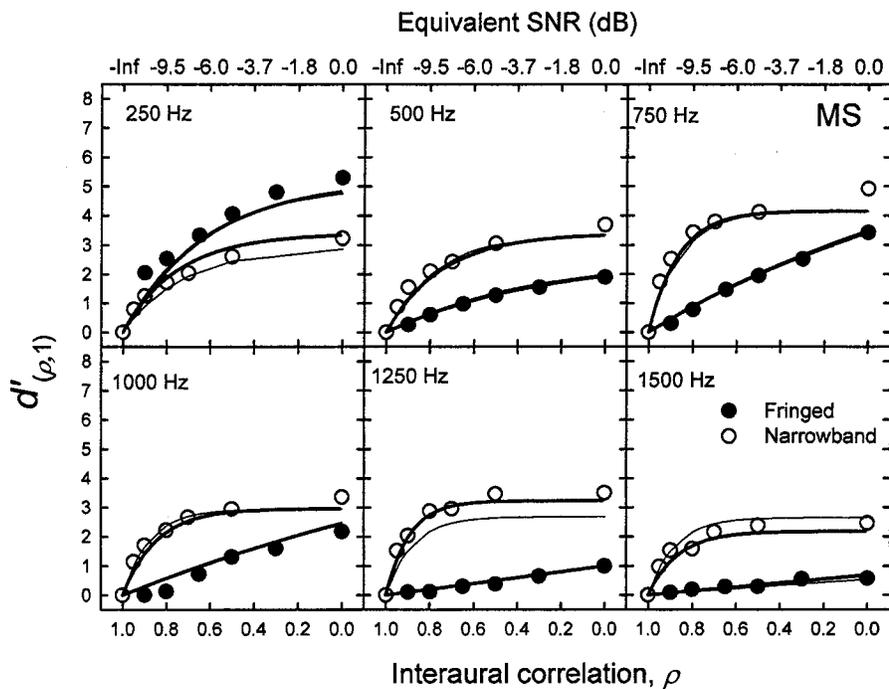


FIG. 5. As in Fig. 4, but for listener MS.

TABLE II. $d'_{(0,1)}$ for each listener in each condition and at each center frequency, fitted separately to the raw data for each listener, condition, and frequency using Eq. (4).

Listener	Fringed center frequency (Hz)						Narrow-band center frequency (Hz)					
	250	500	750	1000	1250	1500	250	500	750	1000	1250	1500
HF	1.26	2.96	2.13	1.43	0.88	0.35	0.75	1.62	1.46	0.51	1.25	0.65
KL	4.09	5.48	3.97	3.68	3.01	1.02	1.46	1.37	1.07	1.16	0.45	0.70
KV	3.26	4.48	3.30	1.92	0.80	0.54	1.51	1.45	2.94	2.43	2.03	1.62
RM	6.11	5.26	3.55	2.09	1.46	0.68	0.81	3.06	2.49	0.66	0.46	0.64
SH	3.38	3.84	3.87	5.14	0.94	0.24	1.43	2.30	3.38	1.37	1.48	1.71
AM	4.22	4.35	2.90	2.45	2.18	1.17	4.65	5.09	3.13	2.65	2.58	1.39
JC							6.26	6.26	6.71	4.78	3.15	3.17
MS	5.29	1.89	3.42	2.17	0.99	0.57	3.22	3.68	4.92	3.35	3.49	2.46
MT	6.75	7.72	8.07	5.81	4.21	2.99	4.94	4.58	3.68	2.82	3.22	2.93
\bar{x}	4.30	4.50	3.90	3.09	1.81	0.95	2.78	3.26	3.31	2.19	2.01	1.70
σ	1.75	1.75	1.78	1.62	1.24	0.88	2.05	1.76	1.71	1.40	1.17	0.97

cies (Durlach and Colburn, 1978, p. 431). Since the BMLD for broadband maskers asymptotes to around 3 dB at high frequencies, it was thought advisable to use a function which could produce the same sort of behavior. In the event, it was found that these two logistic functions could, in combination, produce a very wide range of surfaces, including surfaces with nonmonotonic changes over frequency (e.g., Fig. 8).

Figures 4 and 5 show fits to $d'_{(\rho_1, \rho_2)}$ values from individual listeners, while Figs. 6–8 show fits to averaged $d'_{(\rho_1, \rho_2)}$ values. These fits turned 90 data points (15 discriminations \times 6 frequencies) into a surface representing $d'_{(\rho, 1)}$ as a continuous function of correlation and frequency. For the fringed data (Fig. 6), the pattern of results was similar across listeners, so the $d'_{(\rho_1, \rho_2)}$ values were averaged across all listeners before making a fit. For the narrow-band data, however, four listeners performed markedly better than the other five. These two groups were averaged and fitted separately (Figs. 7 and 8). The thin curves in Figs. 4 and 5 come from fits to the data from the individual listeners concerned. It can be seen that the fit closely approximates the $d'_{(\rho, 1)}$ curves derived at individual frequencies (thick curves).

Table III shows the parameters that were fitted to the three averaged data sets. Figures 6–8 show surface plots for the growth in $d'_{(\rho, 1)}$ as functions of both frequency and the signal-to-noise ratio required to produce the appropriate correlation when a signal is added to noise in the NoS π configuration. The “equivalent signal-to-noise ratio” was calculated using Eq. (7), adapted from Jain *et al.* [1991, Eq. (1)]. Figure 6 shows this function for the averaged data from the fringed condition. Figure 7 shows the function for the four more sensitive listeners in the narrowband condition. Figure 8 shows the same function for the five less sensitive listeners:

$$\text{SNR} = 10 \log_{10} \left(\frac{1 - \rho}{1 + \rho} \right). \quad (7)$$

III. DISCUSSION

The principal interest of the study was to find out how well listeners can discriminate different degrees of interaural correlation across a range of different correlations, rather than just at correlations close to one. The motivation for the study was to further the understanding of binaural masking

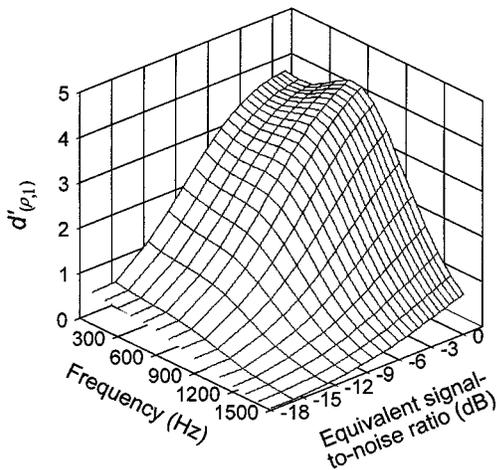


FIG. 6. Surface plot of the growth of $d'_{(\rho, 1)}$ as a function of equivalent SNR and frequency for the *fringed* data. Surface lines are drawn at intervals of 100 Hz in the frequency dimension and 0.05 in the correlation dimension.

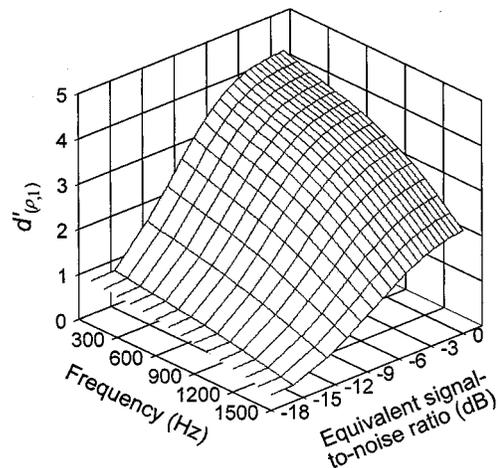


FIG. 7. As in Fig. 6, but plotted for the four *more sensitive* listeners in the *narrow-band* condition.

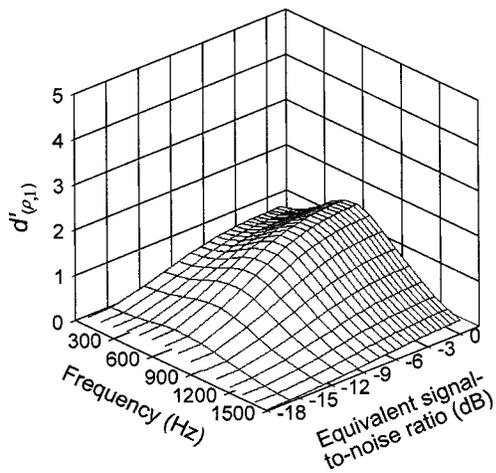


FIG. 8. As in Fig. 7, but plotted for the four *less sensitive* listeners in the *narrow-band* condition.

release and its application to speech recognition. The two main conditions of the experiment were consequently modeled upon the two most common types of BMLD experiment, in which a tonal signal is presented interaurally out-of-phase against either a diotic broadband noise or a diotic narrow-band noise.

A. Relationship to previous discrimination and BMLD experiments

If, as a number of authors have argued, binaural masking release is mediated by the detection of interaural decorrelation, the present stimuli should provide similar cues to the corresponding BMLD stimuli, since the tone in a BMLD stimulus usually acts to decorrelate the noise in its immediate frequency region. The fringed condition, in which listeners discriminated different interaural correlations of a target band of noise embedded within a broadband diotic noise, corresponds to BMLD conditions with a broadband masker. The narrow-band condition, in which only the target band was present, corresponds to BMLD conditions with a narrow-band masker. Consistent with these ideas, the listen-

TABLE III. The fitted logistic parameters that control the values of n and k across frequency to produce an optimum fit with the 90 raw data points in each condition (15 measurements at six frequencies). These parameters determine the surfaces plotted in Figs. 6–8.

$d'_{(\rho,1)}$ parameter	Logistic parameter			
	r	s	t	a
Fringed (mean data)				
k	4.68	0.0027	666	0.023
n	3.17	-0.0047	560	-2.75
Narrow band (4 more sensitive listeners)				
k	6.67	0.0010	1500	0.010
n	2.25	-0.0020	708	-4.37
Narrow band (5 less sensitive listeners)				
k	5.97	0.0021	1500	0.018
n	1.81	-0.0062	423	-5.92

ers' experience of the task was quite different in the two conditions and closely resembled the experience of listening to the corresponding BMLD stimuli.

For the broadband stimuli, listeners heard a “warbling tone” or “whistling noise” standing out from the background in much the same way as listeners hear the target tone standing out from the noise in a broadband-BMLD stimulus. The whistling sound was louder the less correlated the target band was. For the narrow-band stimuli, listeners heard a broadening of the sound image for stimuli of lower correlation, just as listeners perceive a broadened sound image when a tone is added out-of-phase to a diotic narrow-band noise. The image was broader the less correlated the noise band was. Jain *et al.* (1991) performed a similar experiment, although with narrower target bands and only with correlations close to one. They compared the just-noticeable-differences (jnd's) in correlation (from one) of their 10-Hz-wide target bands with the differences in correlation at detection threshold for corresponding BMLD stimuli. They found these values to be very similar, indicating that detection of the decorrelation was sufficient to explain listeners' ability to detect the tones in the NoS π condition of a BMLD experiment.

Taking threshold as $d'_{(\rho,1)} = 1$, and rearranging Eq. (4), one obtains the following equation for the jnd from unity correlation where n and k are the fitting parameters discussed earlier:

$$\text{jnd}_1 = \frac{-\ln(1 - e^{-(k+n)})}{k}. \quad (8)$$

From this equation, jnd's at 500 Hz are calculated to be 0.09 for the fringed condition, 0.06 for the four more sensitive listeners in the narrow-band condition, and 0.13 for the five less sensitive listeners in the narrow-band condition. These thresholds are substantially higher than those reported previously (Pollack and Trittipoe, 1959a, b; Gabriel and Colburn, 1981; Durlach *et al.*, 1986; Koehnke *et al.*, 1986; Jain *et al.*, 1991; Bernstein and Trahiotis, 1992, 1996a, b). It is difficult to determine the reasons for the poorer sensitivity observed here, since there are a large number of differences between the current experimental design and the methods used by other authors. The listeners were probably less well trained and had to contend with stimuli within a session which varied much more widely in correlation than in experiments that only measured jnd's close to a correlation of 1.0. Another factor is the use of batteries of 50 stimulus pairs with a fixed pair of correlation values. Such batteries may be relatively difficult for listeners to optimize their performance on since they may become disheartened when contending with a battery they find difficult, and may find it hard to maintain concentration on a battery they find very easy. The design of our experiments attempted to minimize such effects by testing pairs of correlation values that yielded values of d' in the 0.5–2 area for most listeners; nonetheless our experiments appear to underestimate sensitivity when the results are compared to measurements using other methods.

The present experiment measured sensitivity to differences in correlation as a function of reference correlation, rather than just at a correlation of one. The only published

precedents for this approach are the experiments of Pollack and Trittipoe (1959a, b), who decorrelated broadband, high-pass, or low-pass noise. Although they showed that listeners were sensitive to differences in correlation across a range of correlation values, their stimuli were not similar to those used in BMLD experiments. When the entire spectrum of a noise is decorrelated, listeners tend to hear a broadening of the sound image (as with narrow-band stimuli). Nonetheless, their results were similar to those observed here, in that listeners were found to be very sensitive to differences in correlation close to one and less sensitive at discriminating lower values of correlation. It is interesting to note that while listeners varied widely in their ability to discriminate the correlations of narrow-band stimuli, they all gave very similar patterns for the fringed stimuli. This difference in listener variability also seems to correspond with BMLD data; Bernstein *et al.* (1998) found that some listeners were much better than others at narrow-band BMLD experiments, while performance in broadband noise was relatively consistent across listeners.

B. Relationship to the binaural intelligibility level difference

The binaural intelligibility level difference is very similar to the BMLD, except that listeners are required to identify speech sounds and words, rather than to detect tones. However, a distinguishing feature of this task is that speech sounds are broadband and, when added in the NoS π configuration to noise, they decorrelate all frequency channels to some extent. In order for listeners to make good use of their binaural systems, therefore, it is important that the encoding of embedded signals is graded. That is to say, that the perceptual salience of a speech component, recovered from noise by the binaural system, needs to grow progressively as the speech component becomes more intense. If this were not the case, the system might detect the presence of the speech, but be unable to discern its spectral profile.

The data of Pollack and Trittipoe show that listeners are far more sensitive to changes in correlation near to a correlation of unity than to changes at lower levels of correlation—a highly nonlinear relationship. The narrow-band conditions of the present experiment gave data that were consistent with this finding. However, in the fringed condition, we investigated listeners' ability to make such discriminations in a situation more similar to the understanding of speech in noise, where each subband will display a different interaural correlation. When listeners were required to discriminate different levels of correlation in one subband of a broadband sound, they not only heard the decorrelation in a different way (as a separate sound) but they were more sensitive to changes in correlation at low reference values, than when, as in the narrow-band condition, the whole stimulus was decorrelated. A novel finding is that, while the cumulative d' function in the narrow-band condition is curved at low frequencies, indicating greater sensitivity close to a correlation of one, the function straightens out to give a near-linear relationship at 1 kHz between interaural correlation and cumulative d' (Figs. 4 and 5). Further, when correlation is reexpressed as the equivalent SNR in dB, the relationship

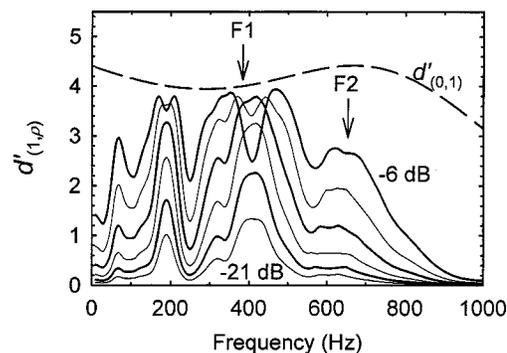


FIG. 9. Perceptually scaled spectra predicted to be recovered by the binaural system for the synthesized vowel /ɔ/ as in “bored” (British English) at overall SNRs from -6 to -21 dB in 3-dB steps. The vowel was synthesized by a Klatt (1980) cascade formant synthesizer ($F1=385$ Hz, $F2=657$ Hz) and added to speech-shaped noise in the NoS π configuration. Frequency selectivity was modeled using the Patterson *et al.* (1987, 1988) gammatone filterbank. Corresponding frequency channels from the left- and right-hand channels were cross correlated in the range ± 2 ms. The maximum product-moment correlation was converted into $d'_{(p,1)}$ using Eqs. (4)–(6) and the parameters derived from the fringed stimuli in Table III. The dashed line shows the maximum cumulative sensitivity $d'_{(0,1)}$ as a function of frequency derived from Eqs. (4)–(6).

between $d'_{(p,1)}$ and SNR is quite linear for equivalent signal-to-noise ratios between about -18 and -6 dB even at low frequencies and in both the fringed and narrow-band conditions (Figs. 6 and 7). Thus, a fixed increment in d' corresponds to a fixed increment in dB (equivalent SNR) over a range of masked-signal intensities. In other words, where the levels of embedded signals are discriminated by virtue of the different degrees of decorrelation they generate, Weber’s law appears to hold, at least approximately.

The results of the fringed condition therefore support the idea that listeners may be able to use the spectral profile of interaural correlation as an index of the intensity of otherwise masked components of speech. Figure 9 illustrates the implications of these data for the internal representation of speech, as provided by the binaural system. The figure shows a recovered spectrum for the synthesized vowel /ɔ/ (as in “bored”) embedded in speech-shaped masking noise in the NoS π binaural configuration. The vowel was synthesized with the Klatt (1980) cascade vowel synthesizer at a fundamental frequency of 100 Hz and with first and second formants at 385 and 657 Hz, respectively. This vowel was selected for its low second formant; the second formants of many vowels would be outside the range of frequencies for which the binaural system is most effective. Interaural coherence (the maximum interaural cross correlation calculated in the range ± 2 ms) was measured in corresponding frequency channels of a stereo pair of gammatone filterbanks (Patterson *et al.*, 1987, 1988), and these values were converted to $d'_{(p,1)}$, using Eqs. (4)–(6). The parameters employed (Table III) were from fitting the pooled results of the fringed condition.

Figure 9 shows a roughly linear growth of peaks in $d'_{(p,1)}$ corresponding to spectral features with increasing SNR. Spectral peaks attributable to both harmonics and formant peaks are visible at a succession of SNRs. The most promi-

nent peak, especially at the lower SNRs, is for the fourth harmonic, just above $F1$.

C. Relationship to dichotic pitches and other binaural phenomena

The parameters for fringed stimuli in Table III can be used in combination with Eqs. (4)–(6) in order to generate predictions of the binaurally recovered spectrum for any binaural stimulus. For instance, these parameters and equations were used by Culling (2000) in order to generate the perceptually-scaled binaurally-recovered spectra for various Fourcin-pitch stimuli in his “revised” mE-C model (Culling, 2000, Figs. 6 and 7). This method of prediction can be applied to any dichotic pitch stimulus in order to determine whether the correct pitch can be predicted from the pattern of interaural coherence across frequency.

IV. CONCLUSION

The results of the present experiment indicate that listeners can discriminate different levels of interaural correlation, especially in the context of a broadband (fringed) stimulus. It is possible that the information provided by spectral variation in correlation may account for listeners’ improved understanding of the speech when listening to the NoS π binaural configuration compared to the NoSo configuration.

ACKNOWLEDGMENTS

This work was supported by the MRC and NIDCD Grant No. DC00100.

APPENDIX: THE GRAM–SCHMIDT PROCEDURE

The Gram–Schmidt procedure allows one to start with one set of functions and to generate a second set of functions that are pairwise orthogonal (uncorrelated with each other) and normalized to have equal energy. This new set of functions is useful in several ways: all the original functions can be written as linear combinations of the new functions; all of the new functions have the same energy; all the new functions are uncorrelated.

The following procedure is simplified here to the case with only two waveforms in the set, because that is all we need. The procedure can be generalized to any number of waveforms. Also, we present the case of discrete-time (sampled) waveforms. In Wozencraft and Jacobs (1965), the case of continuous-time waveforms is presented.

Assume two N -sample nonzero waveforms a and b , that are not perfectly correlated with each other, such as two independent samples of noise. Represent them as $\{a_i\}$ and $\{b_i\}$ for $i = 1, \dots, N$. There are four steps, outlined next, that result in two orthogonal waveforms $\{a_i\}$ and $\{b'_i\}$, where $\{b'_i\}$ has identical rms power to, and zero correlation with, $\{a_i\}$. Following this orthogonalization procedure, a and b' can be mixed according to Eq. (A1), to give a precise correlation, ρ , between a and the mixture m :

$$m_i = \rho a_i + \sqrt{1 - \rho^2} b'_i \quad \text{for } i = 1, \dots, N. \quad (\text{A1})$$

The Gram–Schmidt orthogonalization procedure has four steps as follows:

- (1) Calculate the rms power of a and b :

$$a_{\text{rms}} = \sqrt{\frac{\sum_{i=1}^N a_i^2}{N}}, \quad (\text{A2})$$

$$b_{\text{rms}} = \sqrt{\frac{\sum_{i=1}^N b_i^2}{N}}. \quad (\text{A3})$$

- (2) Calculate the correlation, ρ_{ab} , between a and b :

$$\rho_{ab} = \frac{\sum_{i=1}^N a_i b_i}{a_{\text{rms}} b_{\text{rms}}}. \quad (\text{A4})$$

- (3) Subtract the correlated component of a from a scaled version of b . This subtraction yields c , which has zero correlation with a :

$$c_i = \frac{a_{\text{rms}}}{b_{\text{rms}}} b_i - \rho_{ab} a_i \quad \text{for } i = 1, \dots, N. \quad (\text{A5})$$

- (4) Scale c to get b' which has zero correlation with, and equal power to, a :

$$b'_i = \frac{c_i}{\sqrt{1 - \rho_{ab}^2}}. \quad (\text{A6})$$

The entire process of generating b' may be summarized as

$$b'_i = \frac{a_{\text{rms}}}{b_{\text{rms}} \sqrt{1 - \rho_{ab}^2}} b_i - \frac{\rho_{ab}}{\sqrt{1 - \rho_{ab}^2}} a_i \quad \text{for } i = 1, \dots, N. \quad (\text{A7})$$

¹Caution should be observed in reading Pollack and Trittipoe’s paper, since the authors did not calculate their correlation values correctly (Jeffress and Robinson, 1962). A corrected version of some of their data may be found in Durlach *et al.* (1986, Table II).

²When fits were made for all frequencies simultaneously, the search space was found to contain many local minima, often producing similar surfaces by using quite different parameter sets. It was, therefore, found necessary to restart the “simplex” search algorithm repeatedly with randomly perturbed starting parameters in order to guarantee an optimal fit.

- Akeroyd, M. A., and Summerfield, A. Q. (1999). “A binaural analog of gap detection,” *J. Acoust. Soc. Am.* **105**, 2807–2820.
- Baker, R. J., Rosen, S., and Darling, A. M. (1998). “An efficient characterization of human auditory filtering across level and frequency that is also physiologically reasonable,” in *Psychological and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 81–88.
- Bernstein, L. R., and Trahiotis, C. (1992). “Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies,” *J. Acoust. Soc. Am.* **91**, 306–316.
- Bernstein, L. R., and Trahiotis, C. (1996a). “On the use of the normalized correlation as an index of interaural envelope correlation,” *J. Acoust. Soc. Am.* **100**, 1754–1763.
- Bernstein, L. R., and Trahiotis, C. (1996b). “The normalized correlation: Accounting for binaural detection across center frequency,” *J. Acoust. Soc. Am.* **100**, 306–316.
- Bernstein, L. R., Trahiotis, C., and Hyde, E. L. (1998). “Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broadband noise,” *J. Acoust. Soc. Am.* **103**, 2069–2078.
- Bronkhorst, A. W., and Plomp, R. (1988). “The effect of head-induced interaural time and level differences on speech intelligibility in noise,” *J. Acoust. Soc. Am.* **83**, 1508–1516.

- Carhart, R., Tillman, T. W., and Greetis, K. R. (1969a). "Release from multiple maskers: Effects of interaural time disparities," *J. Acoust. Soc. Am.* **45**, 411–418.
- Carhart, R., Tillman, T. W., and Greetis, K. R. (1969b). "Perceptual masking in multiple sound backgrounds," *J. Acoust. Soc. Am.* **45**, 694–703.
- Colburn, H. S. (1996). "Computational models of binaural processing," in *Auditory Computation*, edited by H. Hawkins, T. McMullen, A. Popper, and R. Fay (Springer, New York).
- Culling, J. F., and Summerfield, Q. (1995). "Perceptual segregation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," *J. Acoust. Soc. Am.* **98**, 785–797.
- Culling, J. F. (2000). "Dichotic pitches as illusions of binaural unmasking. III. The existence regions of the Fourcin pitch," *J. Acoust. Soc. Am.* **107**, 2201–2208.
- Durlach, N. I., and Colburn, H. S. (1978). "Binaural phenomena," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. IV.
- Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (1986). "Interaural correlation discrimination. II. Relation to binaural unmasking," *J. Acoust. Soc. Am.* **79**, 1548–1557.
- Gabriel, K. J., and Colburn, H. S. (1981). "Interaural correlation discrimination. I. Bandwidth and level dependence," *J. Acoust. Soc. Am.* **69**, 1394–1401.
- Grantham, D. W. (1982). "Detectability of time-varying interaural correlation in narrow-band noise stimuli," *J. Acoust. Soc. Am.* **72**, 1178–1184.
- Jain, M., Gallagher, D. T., Koehnke, J., and Colburn, H. S. (1991). "Fringed correlation discrimination and binaural detection," *J. Acoust. Soc. Am.* **90**, 1918–1926.
- Jeffress, L. A., and Robinson, D. E. (1962). "Formulas for the coefficient of correlation for noise," *J. Acoust. Soc. Am.* **34**, 1658–1659.
- Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 838–844.
- Koehnke, J., Colburn, H. S., and Durlach, N. I. (1986). "Performance in several binaural-interaction experiments," *J. Acoust. Soc. Am.* **79**, 1558–1562.
- Levitt, H., and Rabiner, L. R. (1967a). "Binaural release from masking for speech and gain in intelligibility," *J. Acoust. Soc. Am.* **42**, 601–608.
- Levitt, H., and Rabiner, L. R. (1967b). "Predicting binaural gain in intelligibility and release from masking for speech," *J. Acoust. Soc. Am.* **42**, 820–829.
- Licklider, J. C. R. (1948). "The influence of interaural phase relations upon the masking of speech by white noise," *J. Acoust. Soc. Am.* **20**, 150–159.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1987). "An efficient auditory filterbank based on the gammatone function," paper presented to the IOC speech group on auditory modelling at the Royal Signal Research Establishment, 14–15 Dec.
- Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1988). "Spiral vos final report, Part A: The auditory filter bank," Cambridge Electronic Design, Contract Report (APU 2341).
- Pollack, I., and Trittipoe, W. J. (1959a). "Binaural listening and interaural noise cross correlation," *J. Acoust. Soc. Am.* **31**, 1250–1252.
- Pollack, I., and Trittipoe, W. J. (1959b). "Interaural noise correlation: Examination of variables," *J. Acoust. Soc. Am.* **31**, 1616–1618.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988). *Numerical Recipes in C* (Cambridge U. P., Cambridge).
- Wozencraft, J. M., and Jacobs, I. M. (1965). *Principles of Communication Engineering* (Wiley, New York).